



#3

By Express Mail # EL628565525US

**APPLICATION FOR UNITED STATES  
LETTERS PATENT**

**A TRANSPARENT STEREO WIDENING ALGORITHM FOR LOUDSPEAKERS**

Inventor:

**Ole KIRKEBY**



## **BACKGROUND OF THE INVENTION**

### **1. Field of the Invention**

This invention relates to spatially extending a sound stage beyond the positions of two loudspeakers for enhanced enjoyment of two-channel stereo recordings.

### **2. Description of the Related Art**

The music that has been recorded over the last four decades is almost exclusively made in the two-channel stereo format which consists of two independent tracks, one for a left channel L and another for a right channel R. The two tracks are intended for playback over two loudspeakers, and they are mixed to provide a desired spatial impression to a listener positioned centrally in front of two loudspeakers that ideally span 60 degrees (*i.e.* relative to the vantage point of the listener, the loudspeakers are at angles of  $\pm 30$  degrees). A limited spatial impression can also be experienced from other listening positions. The two-channel stereo format is also used for the final delivery of many other types of entertainment audio, such as MPEG-2 digital television broadcasts with multiple digital sound channels, digital versatile discs (DVDs), videotapes, CD's, audiocassettes, and video games.

In many situations, it is advantageous to be able to modify the inputs to the two loudspeakers in such a way that the listener perceives the sound stage as extending beyond the positions of the loudspeakers at both sides. This is particularly useful when a listener wants to play back a stereo recording over two loudspeakers that are positioned quite close to each other. The loudspeakers contained in a stereo television, for example, or positioned on either



side of a computer monitor usually span significantly less than the recommended 60 degrees. Nevertheless, a widening of the sound stage is generally perceived as a pleasant effect regardless of the position of the loudspeakers, and many stereo widening schemes have been developed for this task over the years.

5           It is well known that when the polarity of one of the two loudspeakers in a conventional stereo setup is reversed, the sound stage becomes blurred in a way which is generally perceived to be undesirable. Nevertheless, this phenomenon demonstrates that it is possible to achieve a spatial effect simply by feeding the two loudspeakers with two coherent signals that are out of phase. It can be shown that at very low frequencies the signals fed to  
10   the two loudspeakers must be almost exactly out of phase in order to make the sound stage extend beyond the loudspeakers [Kirkeby et al., *Virtual Source Imaging using the Stereo Dipole*, the 103<sup>rd</sup> Convention of the Audio Engineering Society in New York, September 26-29, 1997, AES preprint no. 4574-J10].

          A stereo widening processing scheme generally works by introducing cross-talk  
15   from the left input to the right loudspeaker, and from the right input to the left loudspeaker. The audio signal transmitted along direct paths from the left input to the left loudspeaker and from the right input to the right loudspeaker are usually also modified before being output from the left and right loudspeakers.

          As described in U.S. Patent Nos. 4,748,669 and 5,412,731, sum-difference  
20   processors can be used as a stereo widening processing scheme mainly by boosting a part of the difference signal,  $L - R$ , in order to make the extreme left and right part of the sound



stage appear more prominent. Consequently, sum-difference processors do not provide high spatial fidelity since they tend to weaken the center image considerably. They are very easy to implement, however, since they do not rely on accurate frequency selectivity. Some simple sum-difference processors can even be implemented with analogue electronics without the need for digital signal processing.

Another type of stereo widening processing scheme is an inversion-based implementation, which generally comes in two disguises: cross-talk cancellation networks and virtual source imaging systems. A good cross-talk cancellation system can make a listener hear sound in one ear while there is silence at the other ear whereas a good virtual source imaging system can make a listener hear a sound coming from a position somewhere in space at a certain distance away from the listener. Both types of systems essentially work by reproducing the right sound pressures at the listener's ears, and in order to be able to control the sound pressures at the listener's ears it is necessary to know the effect of the presence of a human listener on the incoming sound waves. U.S. Patent No. 3,236,949 discloses the inversion-based implementations by designing a simple cross-talk cancellation network based on a free-field model in which there are no appreciable effects on sound propagation from obstacles, boundaries, or reflecting surfaces. Later implementations use sophisticated digital filter design methods that can also compensate for the influence of the listener's head, torso and pinna (outer ear) on the incoming sound waves. See *e.g.* U.S. Patent Nos. 4,975,954, 5,666,425, 5,727,066, 5,862,227, 5,917,916.

As an alternative to the rigorous filter design techniques that are usually required for an inversion-based implementation, U.S. Patent No. 5,046,097 derives a suitable set of filters from experiments and empirical knowledge. This implementation is therefore based on tables whose contents are the result of listening tests.

5 It is common to all the implementations mentioned above that they process a substantial part of the audio frequency range. U.S. Patent No. 4,975,954 restricts the processing to affect only frequencies below 10kHz, Gardner suggests the processing cut-off to be at 6kHz [W.G. Gardner, *3-D Audio Using Loudspeakers*, Kluwer Academic Publishers, 1998, pp. 68-78], and it is mentioned that the techniques described in U.S. Patent No. 10 5,046,097 still work even if the processing is restricted to affect frequencies between 200Hz and 7kHz only. Ward and Elko [S. L. Gay and J. Benesty (Editors), *Acoustic Signal Processing for Telecommunication*, pp. 313-317 of Chapter 14, Kluwer Academic Publishers, 2000] suggests splitting up the processing into four different frequency bands: low ( $< 500\text{Hz}$ ), low-mid ( $500\text{Hz} < f < 1.5\text{kHz}$ ), high-mid ( $1.5\text{kHz} < f < 5\text{kHz}$ ), and high ( $> 5\text{kHz}$ ). Only mid 15 frequencies are processed ( $500\text{Hz} < f < 5\text{kHz}$ ) but it is necessary to use four loudspeakers for the reproduction, two closely spaced ( $\pm 7$  degrees recommended) and two widely spaced ( $\pm 30$  degrees recommended).

The widening of the sound stage usually comes at a price. It is difficult to achieve a convincing spatial effect without introducing spectral coloration (*i.e.* certain parts of 20 sound spectrum become more emphasized versus other parts of the sound spectrum) of the original recording. Reflections from the acoustic environment, such as the walls and furniture

in an ordinary living room, tend to make this undesirable spectral coloration effect even more noticeable. Consequently, a stereo widening processing scheme often degrades the quality of the original recording, particularly at positions away from the "sweet spot" (the optimal listening position for which the stereo widening scheme is designed). At non-ideal listening positions, which may be only a matter of centimeters away from the sweet spot, the processing provides the listener with little or no spatial effect but the spectral coloration is noticeable in all of these non-ideal listening positions. Ideally though, a listener who is not in the sweet spot should not be able to tell whether the processing is "on" or "off". It would therefore be advantageous to have a transparent stereo widening algorithm for loudspeakers that maximizes the spatial effect for a listener sitting in the sweet spot while preserving the quality of the original recording.



## SUMMARY OF THE INVENTION

It is an object of the present invention to provide a system and method of extending the sound stage of two closely spaced loudspeakers without deleteriously affecting the sound quality of the audio signal.

5 In accordance with a first embodiment of the present invention, an audio system is provided for spatially widening a stereophonic sound stage provided by at least two loudspeakers without introducing substantial spectral coloration effects. The audio system comprises (a) a pair of left and right loudspeakers to provide a stereophonic audio output, the left and right loudspeakers being spaced apart from one another; (b) a left channel audio input for inputting a  
10 left channel of an audio signal from an audio source to the left loudspeaker over a first direct signal path; (c) a right channel audio input for inputting a right channel of an audio signal from the audio source to the right loudspeaker over a second direct signal path; (d) a first filter stage along the first direct signal path intermediate the left channel audio input and the left loudspeaker for introducing a delay, which is possibly frequency-dependent, to the left channel of the audio  
15 signal before the left channel is output at the left loudspeaker; (e) a second filter stage along the second direct signal path intermediate the right channel audio input and the right loudspeaker for introducing the delay, which is possibly frequency-dependent, to the right channel of the audio signal before the right channel is output at the right loudspeaker; (f) a third filter stage intermediate the left channel audio input and the right loudspeaker along a first indirect signal path  
20 for adding a first low frequency cross-talk signal at frequencies below approximately 2 kHz derived from the left channel audio input to the delayed right channel of the audio signal; and (g)

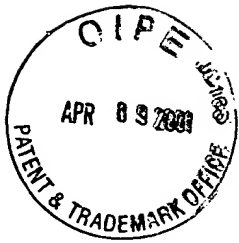
a fourth filter stage intermediate the right channel audio input and the left loudspeaker along a second indirect signal path for adding a second low frequency cross-talk signal at frequencies below approximately 2 kHz derived from the right channel audio input to the delayed left channel of the audio signal. The third and fourth filter stages may each comprise an element for introducing a gain whose absolute value is smaller than approximately 1.0, and a filter having a magnitude response that is not greater than the magnitude response of the first and second first stages at a frequency below approximately 2kHz and that is substantially zero at and above approximately 2kHz. The third and fourth filter stages may also comprise a second element for introducing a second delay that may be greater than the first delay introduced at the first and second filter stages, where the second delay is desired and is not provided by the filter. In one embodiment, the absolute value of the gain of the third and fourth filter stages is between approximately 0.5 and 1.0, and the second delay is between approximately 0 ms and approximately 0.5 ms at frequencies below approximately 2kHz.

In accordance with a second embodiment of the invention, a method is provided for processing an audio signal for reproducing the audio signal as stereophonic sound by at least right and left loudspeakers in a manner that gives an impression that at least part of the sound emanates from a virtual location spaced apart from the actual location of the loudspeakers without introducing a substantial spectral coloration effect. The method comprises (a) inputting an audio signal comprising left and right audio channels to an audio system comprising left and right loudspeakers; (b) filtering the left audio channel at a first filter stage intermediate a left audio channel input and the left loudspeaker along a first direct signal path between the left audio



channel input and the left loudspeaker to delay the left audio channel; (c) filtering the right audio channel at a second filter stage intermediate a right audio channel input and the right loudspeaker along a second direct signal path between the right audio channel input and the right loudspeaker to delay the right audio channel; (d) filtering the left audio channel at a third filter stage intermediate the left channel audio input and the right loudspeaker to add a first low frequency cross-talk at frequencies below approximately 2kHz derived from the left channel audio input to the delayed right channel of the audio signal; and (e) filtering the right audio channel at a fourth filter stage intermediate the right channel audio input and the left loudspeaker to add a second low frequency cross-talk at frequencies below approximately 2kHz derived from the right channel audio input to the delayed left channel of the audio signal. The delayed right audio channel that is added to the first low frequency cross-talk is reproduced at the right loudspeaker, and the delayed left audio channel added to the second low frequency cross-talk is reproduced at the left loudspeaker.

Other objects and features of the present invention will become apparent from the following detailed description considered in conjunction with the accompanying drawings. It is to be understood, however, that the drawings are designed solely for purposes of illustration and not as a definition of the limits of the invention, for which reference should be made to the appended claims. It should be further understood that the drawings are not necessarily drawn to scale and that, unless otherwise indicated, they are merely intended to conceptually illustrate the structures and procedures described herein.



### BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings:

FIG. 1 illustrates the general structure of a stereo widening network, including filters  $H_d$  and  $H_x$  for loudspeakers according to one embodiment of the invention;

5           FIG. 2A illustrates an example of appropriate response characteristics of a filter  $H_d$  that can be used in a direct path between an audio channel input and its corresponding loudspeaker for each of the right and left channels and corresponding loudspeakers;

          FIG. 2B illustrates an example of appropriate response characteristics of a cross-talk filter  $H_x$  used in an embodiment of the invention to introduce a cross-talk signal from a  
10       first audio channel to a second audio channel;

          FIG. 3A illustrates the components of one embodiment of a cross-talk filter  $H_x$  including a consecutive gain element  $g_x$ , allpass filter  $A_x(z)$ , and filter  $G_x(z)$ ;

          FIG. 3B illustrates a desirable magnitude response characteristics of filter  $G_x(z)$  of FIG. 3A;

15           FIG. 4 illustrates an implementation of the stereo widening network according to one embodiment of the invention using linear phase finite impulse response (FIR) filters; and

          FIG. 5 illustrates an implementation of the stereo widening network according to another embodiment of the invention using cascades of second order infinite impulse response (IIR) filters.



## DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENTS

FIG. 1 shows in block form the general structure of a stereo widening network according to the prior art as well as the present invention. The network, which is generally implemented on a digital signal processor (DSP), comprises left and right loudspeakers 10, 20.

5 A digital audio source 30 has separate audio inputs L and R for left and right channels, respectively. (The sound stage can also be widened by placing an additional set of loudspeakers behind a listener.) The audio source 30 is input as a stream that may comprise a live digital audio signal or a digital audio recording stored in any format and on any media. For example, audio source 30 may be an audio signal stored on a DVD, or in the MP3 format.

10 As another example, audio source 30 may be an audio signal that is a soundtrack to a movie, television, or is part of any multimedia program.

A left channel of audio source 30 is input at left channel input L and a right channel of audio source 30 is input at right channel input R. The left channel is filtered by a filter  $H_d$  40, is added at adder 60 to cross-talk from the right channel that is filtered by filter  $H_x$  60, and is output at left loudspeaker 10. Similarly, the right channel is filtered by a filter  $H_d$  70, is added at adder 90 to cross-talk from the left channel that is filtered by filter  $H_x$  80, and is output from right speaker 20. (It should be noted that term "cross-talk" is used herein to refer to the part of the audio signal that is leaked from one input to the 'opposite' output, rather than to refer, as is common, to the acoustic path from a loudspeaker to the 'opposite' ear of a listener.) Generally, rather than implementing them as a single filter,  $H_d$  and  $H_x$  are each implemented as a filter stage comprising multiple components as is discussed below.

The distinctiveness and advantages of the present invention lies in the derivation and the properties of  $H_d$  and  $H_x$ . The choice of  $H_d$  and  $H_x$  is motivated by the need for achieving a good spatial effect without degrading the quality of the original audio source material. In the present invention,  $H_d$ , used for both filters 40, 70, is a filter with a flat magnitude response, thus leaving the magnitude of the signal input thereto unchanged while introducing a group delay (it should be noted that group delays, and delays can vary as a function of frequency). Thus, significantly,  $H_d$  permits the respective channel from audio source 30 to pass through on a direct path to that channel's respective loudspeaker without any change in magnitude.  $H_x$ , used for both filters 50, 80, is a filter whose magnitude response is substantially zero at and above a frequency of approximately 2kHz, and whose magnitude response is not greater than that of  $H_d$  at any frequency below approximately 2kHz. In addition, a group delay is introduced by filter  $H_x$  that is generally greater than the group delay introduced by filter  $H_d$ .

FIGS. 2A and 2B show examples of appropriate magnitude responses of  $H_d$  and  $H_x$ , respectively, for the present invention. The magnitude response of  $H_x$  is bounded in the vertical direction by the magnitude of  $H_d$ , and in the horizontal direction by approximately 2kHz. The magnitude of frequencies above approximately 2kHz are designed not to be affected by filter  $H_x$  because altering the magnitude of these frequencies above approximately 2kHz creates undesirable spectral coloration.

FIG. 3A illustrates how filter  $H_x$  can be separated into three consecutive components which allow separate control over the magnitude and phase responses: (1) a cross-

talk path gain  $g_x$  whose absolute value is smaller than one, (2) a frequency-independent delay, or frequency-dependent delay introduced for example by an allpass filter  $A_x$  [Regalia et al. *The Digital All-Pass Filter: A Versatile Signal Processing Building Block*", Proceeding of the IEEE, 76(1), pp. 19-37, January 1988] (or  $A_x(z)$  in the z-transform domain), and (3) a filter  $G_x$  ( $G_x(z)$  in the z-transform domain) whose maximum magnitude response is one at frequencies below 2kHz, and is substantially zero at frequencies at and above 2kHz. FIG. 3B shows an example of the magnitude response of filter  $G_x$ . Filter  $A_x$  is an unnecessary element where filter  $G_x$  can provide the desirable delay otherwise provided by filter  $A_x$  (e.g.  $G_x$  is an FIR filter as described below.)

In practice, it has been found that the filter  $H_x$  obtained from the following combination of  $g_x$ ,  $A_x(z)$  and  $G_x(z)$  gives very good results (i.e. the desired stereo widening with minimal spectral coloration):  $g_x \approx -0.8$ ,  $A_x(z)$  is a frequency-independent delay of about 0.2ms (which results in a delay of about 10 samples relative to the delay introduced by  $H_d$  at a sampling frequency of about 48kHz), and  $G_x(z)$  is a bandpass filter that blocks very low frequencies (below approximately 250 Hz) as well as frequencies above approximately 2kHz. The highpass-characteristic of  $G_x(z)$  wherein frequencies below approximately 250 Hz are blocked prevents very low frequencies in one channel of the audio signal from being canceled out by the out-of-phase cross-talk that is added from the other channel. (The left and right channels are 180 degrees out of phase at 0Hz and slightly less out of phase at low frequencies.) Preventing the loss of low frequencies between approximately 0 and approximately 250 Hz ensures that a natural balance is maintained between low and high frequencies. However, the

bandpass characteristic of  $G_x(z)$  might not always be required. If the loudspeakers used for the reproduction are very poor, for example, and they are not capable of emitting any significant sound at low frequencies anyway, then there is no need to process this frequency range at all, and in that case  $G_x(z)$  could be a simple lowpass filter, instead of the filter with a magnitude response shown in FIG. 3B.

When the absolute value of  $g_x$  is smaller than approximately 0.5, the spatial effect of the processing is so subtle that in most situations it will not be beneficial to the listener. When the delay introduced by  $A_x(z)$  is greater than approximately 0.5ms (which results in a delay of approximately 24 samples relative to the delay introduced by  $H_d$  at a sampling frequency of approximately 48kHz), the spatial effect of the processing becomes somewhat unnatural sounding to the human ear (sometimes called "phasiness") and is uncomfortable to listen to, whereas short delays, or even no delay, still has an overall positive effect on the perceived sound. The absolute value of  $g_x$  should therefore be between approximately 0.5 and 1.0, and the group delay function of  $A_x(z)$  relative to the delay introduced by  $H_d$  must be between approximately 0 ms and approximately 0.5 ms at frequencies below about 2kHz. The value of the group delay function of  $A_x(z)$  above approximately 2kHz is irrelevant since those frequencies are blocked by  $G_x(z)$  anyway.

If the sampling frequency is relatively low, the stereo widening algorithm may be conveniently implemented by realizing the cross-talk filters  $H_x$  as a gain  $g_x$  followed by a linear phase finite impulse response (FIR) filter which is used for  $G_x(z)$ , and by realizing the direct-path filters  $H_d$  as the delay of  $z^{-(N-N_x)}$ , as shown in FIG. 4.  $N$  is the group delay of the

linear phase FIR filter, which is of the order of 100 at 48kHz, and scales up and down linearly with the sampling frequency. Thus, for example,  $N$  is of the order of 25 at 12kHz. (No separate group delay source such as  $A_x$  is necessary in this implementation because the delay is added by the FIR filters.) Since the group delay introduced by the linear phase filters are constant as a function of frequency, it is sufficient to insert a delay line in the direct path in order to match the delay of the cross-talk path up to a desired amount of delay, thereby enabling the provision of a controllable amount additional delay in the cross-talk path, relative any delay in the direct path. For example, if the group delay in the cross-talk path is 23 samples at a sampling frequency of approximately 12kHz, then inserting a delay of about 20 samples in the direct path with filter  $H_d$  ensures that the cross-talk path is delayed by about 3 samples, which corresponds to approximately 0.25 ms, relative to the direct path. A fractional delay can be used to match the delays with sufficient accuracy if necessary.

An audio signal having a bandwidth greater than approximately 2kHz, including a signal whose sampling frequency is relatively low (*e.g.* approximately 8 kHz - approximately 12 kHz) or relatively high (*e.g.* approximately 32 kHz - approximately 48 kHz), may be processed by the stereo widening algorithm of the present invention. However, processing at a low sampling frequency does not necessarily mean that the stereo widening algorithm is being used for a lo-fi (low fidelity) application. As an example, where the algorithm is used for processing signals at a low sampling frequency for a hi-fi (high fidelity) application, the audio source signal can be divided into sub-bands. In the simplest case, the audio source signal at whatever frequency it is input can be decomposed into two frequency bands: a base band that

contains energy only at frequencies below approximately 2kHz ( $f < 2\text{kHz}$ ) and a band that contains energy only at frequencies greater than approximately 2 kHz ( $f > 2\text{kHz}$ ). The spatial processing need only be applied to the base band, which makes the processing less expensive than if the entire signal were processed. The main computational expense is in the splitting, and recombining, of the two frequency bands. Perceptual coding schemes, such as MP3, split up the signal into different frequency bands anyway. It is therefore relatively straightforward to combine the perceptual coding with the spatial processing of the lower frequency sub-band as described in a hybrid type of algorithm. Care must be taken to match the delays across the frequency range, though, when the sub-bands are combined to form the final output.

At high sampling rates, the FIR filters necessary for shaping the frequency response of  $G_x(z)$  below 2kHz contain so many coefficients that in most practical applications they are prohibitively expensive to implement. One alternative for cross-talk filter  $H_x$  is to use interpolated FIR (IFIR) filters [as described by Saramäki et al., *Design of Computationally Efficient Interpolated FIR Filters*, IEEE Transactions on Circuits and Systems, 35(1), pp. 70-88, January 1988) and Y. Lin and P.P. Vaidyanathan, *An Iterative Approach to the Design of IFIR Matched Filters*, Proc. IEEE International Symposium on Circuits and Systems, pp. 2268-2271, 1997], which are made up of cascades of dense and sparse FIR filters, but even IFIR filters are sometimes too expensive to implement at the sampling frequencies used for high-quality audio. Both FIR and IFIR implementation are suitable for implementation in 16-bit fixed-point precision.



FIG. 5 shows another implementation of the stereo widening algorithm that is particularly suitable for operating at high sampling frequencies, such as the standard sampling rates of 44.1kHz and 48kHz commonly used for high-quality audio, because it is more economical and efficient at higher frequencies. (It is believed that the IIR filter implementation is more efficient than the FIR filter implementation even at 10 kHz and above.) The IIR implementation uses cascades of substantially identical second order infinite impulse response (IIR) filters that are applied to each of the cross-talk paths. Each cross-talk filter  $H_x$  of FIG. 1 is realized in the implementation of FIG. 5 as a gain  $g_x$  followed by a delay of  $z^{-N}$  and a cascade of at least four filters in each cross-talk path, including a pair of high-pass filters  $H_{hi}(z)$  followed by a pair of low-pass filters  $H_{lo}(z)$ . A frequency-dependent delay can be implemented by replacing  $z^{-N}$  with an allpass filter  $A_x$ .

$z^{-N}$  is the delay intentionally introduced into the cross-talk path relative to the delay in the direct path.  $z^{-N}$  is between approximately 0 and approximately 0.5ms depending on the spacing between the right and left loudspeakers (shorter delays for narrow spacing between loudspeakers 10, 20, longer delays for wider spacing between loudspeakers 10, 20). The delay  $z^{-N}$  is of the order of 10 samples at 48kHz (which is equivalent to 0.2ms), and, as with the delay  $z^{-(N-N_x)}$  in the embodiment of FIG. 4,  $z^{-N}$  also scales up and down linearly with the sampling frequency.

$H_{hi}(z)$  starts cutting on at approximately 250Hz and  $H_{lo}(z)$  starts cutting off at approximately 1.5kHz. This cascade of filters provides a bandpass filter having a magnitude response as shown in FIG. 3B. The doubling of filters  $H_{hi}(z)$  and  $H_{lo}(z)$  in the cross-talk path (*i.e.*

providing them as pairs) squares the magnitude responses of filters. Consequently, in the pass-band, the magnitude response is still 1 but the doubling of filters causes the roll-off to be steeper.

Rather than implementing  $H_x$  in FIG. 5 with four filters, including lowpass filters  $H_{lo}(z)$  and highpass filters  $H_{hi}(z)$ ,  $H_x$  can be implemented as having only the simple lowpass characteristic of FIG. 2B without the highpass characteristic by using a cascade of two filters only, those filters being the pair of lowpass filters  $H_{lo}(z)$  (and omitting the pair of highpass filters  $H_{hi}(z)$ ).

Additionally, in the implementation of FIG. 5, a pair of allpass filters  $A_{hi}(z)$  and  $A_{lo}(z)$  are inserted into each of the direct paths such that the group delays in each of the direct and cross-talk paths are substantially perfectly matched as a function of frequency to the extent desired (and any desired amount of delay  $z^{-N}$  can be controllably and separately inserted into the cross-talk path). The group delay of  $A_{hi}(z)$  is designed to be the same as the group delay introduced by  $H_{hi}(z)^* H_{hi}(z)$  and the group delay of  $A_{lo}(z)$  is designed to be the same as that of  $H_{lo}(z)^* H_{lo}(z)$ . This can be accomplished using well known filter design principles: the magnitude response of filters  $B(z)$ , where  $B(z)$  is  $H_{hi}(z)^* H_{hi}(z)$  or  $H_{lo}(z)^* H_{lo}(z)$ , is shaped to have double poles, and the corresponding allpass filter  $A(z)$ , whether  $A_{hi}(z)$  or  $A_{lo}(z)$ , respectively, compensates for the group delay of  $B(z)$  with an equivalent group delay by replacing half of the poles of filter  $B(z)$  with zeros at their image positions outside the unit circle.  $B(z)$  can have zeros, in addition to poles, but the zeros must not be inside the unit circle; otherwise their mirror poles are outside the unit circle, which would make the corresponding filters  $A(z)$  unstable. In one implementation, the zeros of filter  $B(z)$  are exactly on the unit circle so that their mirror poles fall on top of the zeros, and therefore cancel them out.

As an alternative to the exact matching of the group delays, one can design the filters in the direct paths and the cross-talk paths to achieve the necessary delays by using approximate methods such as group delay equalization and nearly linear phase IIR filters. Careful design using such methods might lead to other efficient and numerically robust implementations based on either FIR or IIR filters, or combinations thereof.

In order to ensure that the effect of the common group delay of direct and cross-talk paths are inaudible, local variations in the group delay between the group delay of the cross-talk path and the direct path as a function of frequency should not exceed approximately 3ms. This estimate is conservative (so that somewhat larger variations in the group delay may be acceptable), and is a safe range for reproducing most types of audio source material with a relatively high fidelity. The total group delay of the cascade of second order IIR filters shown in FIG. 5, which implements the magnitude response of  $G_x$  shown in Fig. 3B, is well within this range of approximately 0 to approximately 3 ms. The cascades of second order IIR filters are sensitive to loss of numerical precision, and are unlikely to perform well in 16-bit fixed-point precision DSP. A 24-bit fixed-point precision, or floating-point, DSP is usually required.

The decision as to whether to choose the implementation of FIG. 4 or FIG. 5 is relatively unimportant if one has a DSP whose sole purpose is to perform spatial processing of audio. The processing efficiency of the IIR filters may be weighed against the lesser complexity of the FIR filter implementation. Ultimately, the implementation chosen will depend on the application.

In summary, the stereo widening system of the present invention is essentially a hybrid of a cross-talk cancellation system and a virtual source imaging system. A cross-talk cancellation system is capable of making one hear sounds close to one's head (like wearing "headphones in a free field") whereas a virtual source imaging system is capable of making one  
5 hear sounds that are a certain distance away. This stereo widening system makes some frequencies appear to be close to the head at the side, some frequencies appear to be close to the loudspeakers, but outside the angle spanned by them, and some frequencies come from the speakers themselves. In practice, the combination of the three effects gives the listener a pleasant impression of spatial widening when used on music so that the natural sound of the  
10 original recording is preserved regardless of the position of the listener and the properties of the acoustic environment of the loudspeakers, while ensuring that the artifacts of the spatial processing are inaudible.

It should be understood that this invention is generally applicable only for use with loudspeakers, as opposed to other types speakers such as headphones, because there is a  
15 natural cross-talk from loudspeakers 10, 20 generated by overlap of sound output from the loudspeakers 10, 20. The cross-talk introduced by filters  $H_d$  and  $H_x$  is in addition to the cross-talk from loudspeakers 10, 20.

The audio system (or the various filter stages thereof) described above may be arranged in a stand alone system or may be arranged (*i.e.* included) in a device that has  
20 functionality in addition to the playing of an audio signal. One such device is, for example, a digital set-top-box (STB), also known as an IRD, Integrated Receiver Decoder, which receives and decodes digital television signals. The digital television signals are usually transmitted as

packets in accordance with the MPEG-2 standard using a digital television broadcast standard, such as Digital Video Broadcasting (DVB) or a similar standard. Some recent set-top boxes have the ability to receive audio/and video information through an Internet connection, realized either through a broadband cable connection or over a digital video broadcast stream. The  
5 audio and video signals are usually output from the set-top box to a standard television set. However, they could also be output to any display device, such as a computer monitor or a video projector.

Other examples of devices that may include the described audio system include a Mobile Display Appliance (MDA) (*i.e.* a portable display product for receiving audio and/or  
10 video either over a wireless broadband connection, for instance connected to the Internet, or from a digital video broadcast, or both), a personal digital assistant (PDA), a mobile phone, portable game devices (*e.g.* Nintendo Game Boy<sup>®</sup>), other consumer electronic products, etc.

Thus, while there have shown and described and pointed out fundamental novel features of the invention as applied to a preferred embodiment thereof, it will be understood  
15 that various omissions and substitutions and changes in the form and details of the devices illustrated, and in their operation, may be made by those skilled in the art without departing from the spirit of the invention. For example, it is expressly intended that all combinations of those elements and/or method steps which perform substantially the same function in substantially the same way to achieve the same results are within the scope of the invention.  
20 Moreover, it should be recognized that structures and/or elements and/or method steps shown and/or described in connection with any disclosed form or embodiment of the invention may be

incorporated in any other disclosed or described or suggested form or embodiment as a general matter of design choice.